CLAIMS:

1. A method for segmenting compound words in an unrestricted natural-language input, the method comprising:

receiving a natural-language input consisting of a plurality of characters;

constructing a set of probabilistic breakpoints in the natural-language input based on probabilistic breakpoint analysis;

identifying a plurality of linkable components by traversal of substrings of the natural-language input delimited by the set of probabilistic breakpoints; and

returning a segmented string consisting of a plurality of linkable components spanning the natural-language input, wherein the segmented string is interpretable as a compound word.

2. The method of Claim 1, further including the step of analyzing a chart of the linkable components in the case that the segmented string cannot be constructed and returning an unsegmented string interpretable as a partial analysis of a compound word.

3. An apparatus for segmenting compound words in a natural-language input, the apparatus comprising:

a startpoint probability matrix;

an endpoint probability matrix;

a probabilistic breakpoint analyzer coupled to the startpoint probability matrix, the endpoint probability matrix and the natural-language input, the probabilistic breakpoint analyzer being operative to generate a breakpoint-annotated input from the natural-language input; and

a probabilistic breakpoint processor coupled to the probabilistic breakpoint analyzer, the probabilistic breakpoint analyzer being operative to generate a segmented string for the compound words in the natural-language input in response to the breakpoint-annotated input.

4. The apparatus of Claim 3, further comprising a word-boundary analyzer coupled to a lexicon and a memory unit, the word-boundary analyzer being operative to generate the startpoint probability matrix and the endpoint probability matrix.

5. The apparatus of Claim 3, wherein the probabilistic breakpoint processor comprises:

a lexicon;

a chart; and

a breakpoint-delimited substring tester coupled to the lexicon and the chart, the substring tester being operative to receive the breakpoint-annotated input and generate a segmented string in response thereto.

6. The apparatus of Claim 3, wherein the probabilistic breakpoint processor is an augmented probabilistic breakpoint processor comprising:

a lexicon;

a chart;

an augmented breakpoint-delimited substring tester coupled to the chart and the lexicon, the substring tester being operative to identify a plurality of linkable components; and

a chart analyzer coupled to the substring tester and the chart, the chart analyzer being operative to generate the segmented string.

7. The apparatus of Claim 6, wherein the augmented breakpoint-delimited substring tester generates one of:

the segmented string; and

a failure signal.

8. The apparatus of Claim 7, wherein the chart analyzer is coupled to receive the failure signal from the augmented breakpoint-delimited substring tester.

9. The apparatus of Claim 3, wherein the apparatus is configured as a computer readable program code run on a computer usable medium.